

Levels of Analysis for the Study of Environmental Health Disparities

Primary authors:

**M. Soobader, PhD
C. Cubbin, PhD
G.C. Gee, PhD**

Secondary authors:

**A. Rosenbaum, MPH, MS
J. Laurenson, MS
D.C. Payne-Sturges, DrPH**

This paper was written for the “Environmental Health Disparities Workshop: Connecting Social and Environmental Factors to Measure and Track Environmental Health Disparities,” under contract to EPA, Contract No. EP-W-04—049, Task Order No. 11

Table of contents

Executive Summary	1
Workshop Questions.....	5
Introduction	7
Multilevel Models.....	11
Why Multilevel Models?	11
Composition and Context (Partitioning the Variance)	12
Multiple Contexts at Different Levels	13
Multiple Contexts at the Same Level	14
Interaction Effects.....	15
Individual Contextual Interaction.....	16
Contextual and Contextual Interaction.....	17
Modeling Contexts and Individuals Over Time.....	17
Repeated Measures of Individuals.....	18
Repeated Measures of Contexts	19
Types of Contextual Variables Used in Multilevel Models	20
Identifying the Relevant Levels	22
Level Specification.....	23
Populations Not Defined by Conventional Levels.....	24
Issues to Be Considered in Level Specification	25
The Research Question	25
Health Outcome	25
Social Definitions vs. Administrative Definitions vs. Intervention Levels	26
Modifiable Area Unit Problem (MAUP)	26
Exposure Period.....	27
Environmental Exposures	27
Data Linkages for Multilevel Models	29
Geocoding	29
Examples of Data Linkages.....	31
Contextual Variables Linked to Geocoded Data.....	31
State Blood Lead Screening Data Linked to Tax Assessor Data	31
State Blood Lead Screening Data Linked to Census Data	32
NHANES Linked to Aerometric Information Retrieval System (AIRS).....	33
Conclusion	34

Bibliography	36
Appendix A: Place and Social Theory.....	39
Appendix B: Census Data and Census Geography.....	44
Census Data.....	44
Most Commonly Used Levels.....	44
Variations of Conventional Levels	47
Appendix C: Inhalation Exposure to Outdoor Air Pollutants	48
Estimating Outdoor Air Pollutant Concentrations	49
Estimating Exposure Concentrations	50
Characteristics of the Pollutant.....	50
Air Quality Research Goals	51
Air Quality Standards	51
Population Exposure Assessment	51
Multiple Media.....	53
Data Sources	53
Policy Level	54
Appendix D: Ambient and Drinking Water Quality	55
Characteristics of the Exposure.....	55
Water Systems	56
Monitoring Research Goals	57

Executive Summary

We present a fundamental approach for achieving health promotion and sustainability by using multilevel techniques to quantify and monitor socioeconomic and racial/ethnic disparities in environmental health. Reducing racial/ethnic and socioeconomic environmental health disparities requires a comprehensive multilevel conceptual and quantitative approach that recognizes the various levels through which environmental health disparities are produced and perpetuated. Multilevel typically refers to the concept of lower-level units contained within higher-level units; e.g., of individuals nested within groups nested within neighborhoods, or industrial facilities nested within communities. Individuals within groups, groups within local contexts, and local contexts within macro contexts may share similar characteristics. Multilevel techniques explicitly model these correlated data where the assumption of independence between observations is violated and conventional OLS techniques are not appropriate. This is in contrast to procedures that attempt to correct for the correlated structure of the data, such as those used in SUDAAN.

We propose a conceptual framework that incorporates the micro level, contained within the local level, which in turn is contained within the macro level, to inform the present discussion. Using lead exposure as an illustrative example, the micro level refers to the most proximate level factors, such as individual body burdens of lead. The local level refers to the immediate context that surrounds the individual, such as the concentrations of lead in the surrounding soil. The macro level refers to both the larger geospatial region that encapsulates the local level (e.g., states) and/or the broader social context (e.g., political climate and laws regarding lead-based paint in housing). Such a conceptual approach underlies multilevel techniques that allows for the consideration of numerous levels

simultaneously, that is, factors that affect health are simultaneously considered as operating at the level of the individual and the level of contexts. We contend that recent research increasingly places primary emphasis on investigating the micro or individual levels, often to the exclusion of the macro level. And while the micro level is indeed important, a major limitation of focusing only on micro-level processes is that the environmental context itself is removed from the line of inquiry. Inattention to the complex interactions between individuals and their environments may lead to inappropriate science, and thus incomplete interventions and policies.

We discuss the utility of multilevel techniques to examine physical and social environmental and individual-level factors to appropriately quantify and improve our understanding of environmental health disparities.

Multilevel modeling approaches can potentially contribute to environmental health research by providing a mathematical modeling approach for:

- Informing environmental policies and examining the impact of existing policies on local contexts and individual exposures.
- Examining a single environmental exposure that may occur through multiple media operating at different levels simultaneously and interacting at different levels.
- Examining multiple exposures operating at different levels simultaneously, potentially accumulating over time, and interacting with each other.
- Examining exposures differentially affecting subgroups of the population and/or geographic areas.
- Examining the fundamental role of social and economic factors and the need to account for all levels through which these mechanisms influence individual exposures, either directly or through their effect on local environments.

We discuss the reasoning and the methodological approach behind multilevel modeling, including differentiating individual and contextual influences on individual outcomes. Environmental studies are typically conducted at a single level, either at the aggregate/ecologic level or the micro/individual level. These studies have been critiqued due to the incorrect inference of the study results. Multilevel models provide the advantage of identifying and differentiating sources of variation at multiple levels, thus assigning variability to the appropriate level. For example, when the same exposure is measured at multiple levels these models allow us to evaluate the relative importance of the exposure at each level. An important feature of multilevel models is that the data need not be hierarchical. That is, contexts do not need to be neatly nested within each other. This is important as exposures commonly occur in contexts that are not hierarchical, but different contexts may occur at the same level. For example, children may be exposed to lead in the neighborhood but also within the school environment, and neighborhoods may not be nested within school districts.

In addition, multilevel models allow us to examine individual and contextual interactions as well as interactions between different levels of context. Modeling an interaction between the individual and the context provides information about the differential effect of context for individual groups. Interactions may also be examined between different levels of context, e.g., providing information on the effect of city expenditure on different types of neighborhoods. Such observations are important for policy development and resource allocation for preventing environmental exposures. Multilevel models also enable us to examine changes over time, including repeated measures of individuals as in panel studies and repeated measures of contexts as in annual statewide surveys. Longitudinal multilevel models are an important component in monitoring environmental health disparities.

Next we address the questions and principles that guide the choice of levels or geographic units in multilevel studies, with worked examples of air pollution and water quality that tackle these issues included as appendices. These include the research question being addressed, the theoretical pathways linking the micro, local, and macro levels, health outcomes and exposures under consideration, data availability, and the administrative or intervention application of the research. Direct policies, synergistic policy effects, sensitivity periods, and mediating mechanisms also influence the definition of the relevant levels.

Finally, the ways in which different data sources can be combined to produce suitable data for multilevel analyses are addressed. Data requirements for multilevel models require, at a minimum, that observations have identifiers that differentiate the contextual setting(s) of each observation. National survey data are now being geocoded and linked to census data to facilitate multilevel analyses. In addition, because of how data are collected, as for environmental exposures, or because of the sampling strategy of national surveys, geographic identifiers are sometimes readily available on some datasets. Note that the appropriateness of the level at which geographic identifiers are available should be evaluated. We provide some examples of how different data sources can be linked to create multilevel data structures, including census data linked to survey data, census and tax assessor data linked to state health department data, and national health survey data linked to environmental exposure data.

Appendices include a review of the social theories of place and provide a description of commonly used census geographic units. Although numerous challenges in multilevel research remain, we call attention to the emerging conceptual and quantitative approaches for assessing the convergence of social, economic, racial/ethnic, and environmental factors in generating and sustaining environmental health disparities.

Workshop Questions

To ensure that all participants get as much value from the workshop as possible, they are asked to answer the following questions prior to the workshop, within the context of their own situation (e.g., an ongoing, planned, or possible disparity analysis).

1. What is the disparity research question or monitoring objective to be addressed?

2. What is the exposure or health(s) outcome?

3. What are the relevant levels to be considered? See Identifying the Relevant Levels, page 22.

4. What are the units at the micro, local, and/or macro level? See Table 1: Examples of Units at different levels, page 24.

5. What are relevant variables to address the research question or monitoring objective, including racial/ethnic, social, economic, and policy variables?

6. What is the multilevel design? If cross-sectional, see Figures 2-4, pages 13-15; if longitudinal, see Figures 5-6, pages 18-19.

7. What data source(s) can be used to address the research question or monitoring objective?

8. Are the units and variables you require available in the data source? If not, how can this be rectified? See Data Linkages for Multilevel Models, page 29.

9. What are the challenges involved in generating the data?

10. What other issues need to be considered to address the research question or monitoring objective?

INTRODUCTION

This paper was written for the “Environmental Health Disparities Workshop: Connecting Social and Environmental Factors to Measure and Track Environmental Health Disparities,” to present fundamental conceptual and analytic approaches for promoting health and sustainability by using multilevel techniques to quantify and monitor socioeconomic and racial/ethnic disparities in environmental health. From a policy perspective, multilevel models can be used to inform policies and examine the impact of existing policies on local contexts, e.g., neighborhoods and individual exposures.

Executive order 12898 on Environmental Justice states that “environmental human health analyses, whenever practicable and appropriate, shall identify multiple and cumulative exposure” [1]. This core theme recognizes that an individual’s health is a complex function of his/her own characteristics, factors in his/her environmental context, and interactions between the individual and his/her context. However, recent research increasingly places primary emphasis on investigating the micro and individual levels or “downstream” factors, often to the exclusion of more macro or “upstream” factors. And while the micro level is indeed important, a major limitation of focusing only on micro-level processes is that the environmental context itself is removed from the line of inquiry. This has led some to argue that environmental scientists and epidemiologists have become “prisoners of the proximate” [2]. Inattention to the complex interactions between individuals and their environments may lead to inappropriate science, and thus incomplete interventions and policies. This trend toward a micro-level research approach has been noted as a barrier towards understanding racial/ethnic disparities in health, because minority populations tend to live in far more hazardous physical and social environments and have lower levels of socioeconomic position (SEP) than do other populations [3, 4].

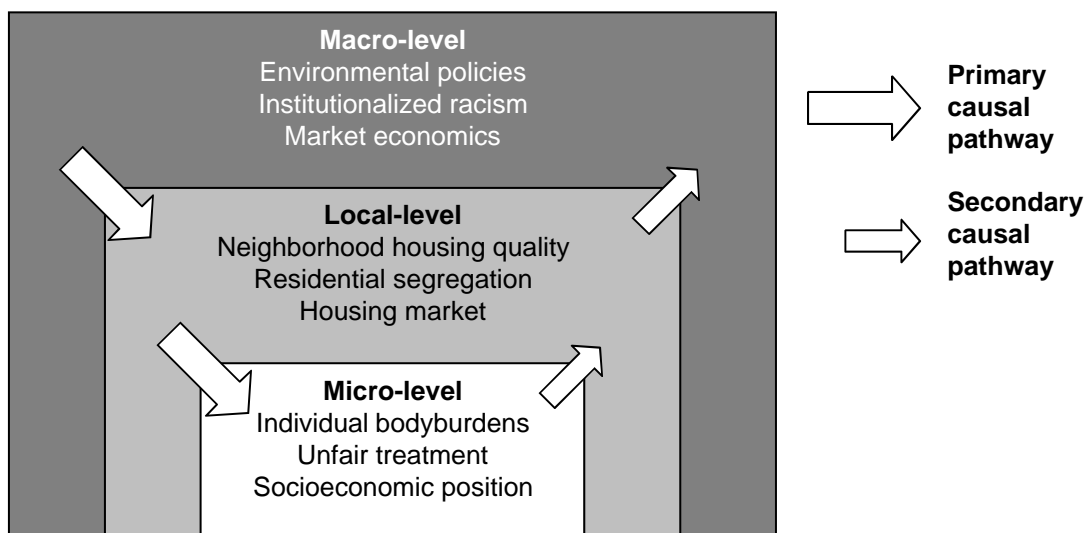
Therefore, reducing racial/ethnic and socioeconomic environmental health disparities requires a comprehensive multilevel research approach that recognizes the various levels through which environmental health disparities are produced and perpetuated. In endorsement of this approach, gene-environment interactions and cumulative risk assessments have also renewed interest in a more complex, multi-factorial and multilevel understanding of health disparities [5]. Gene-environment research suggests that genetic vulnerabilities may be expressed or amplified when certain environmental triggers occur. Cumulative risk assessment examines how multiple exposures from multiple sources over time contribute to health. Recent conceptual environmental frameworks further support this approach. For example, Schulz and Northridge (2004) suggest that fundamental factors, such as the geographic topology and wealth distribution shape intermediate contexts, such as land use and local economies, which in turn shape proximate risks, such as housing quality and/or unfair treatment, ultimately shaping individual health [6]. Similarly, Morello-Frosch and colleagues (2002) suggest that income inequality and social capital influence the ability of local communities to affect environmental and social policy actions, thereby influencing these communities' abilities to resist environmental health stressors such as the siting of hazardous waste facilities and subsequent health effects [7]. Gee and Payne-Sturges (2004) suggest that macro-level residential segregation leads to differential local-level environmental hazards and social stressors, which in turn lead to differential individual-level stressors and subsequent illness and health disparities [8].

Figure 1 presents a generalized conceptual framework of this comprehensive multilevel approach to inform the present discussion. In this framework, the primary causal pathway to a particular health outcome is from the macro to the local to the micro level with the secondary pathway as a feedback loop. Using lead exposure as an illustrative example, the micro level refers to the most proximate-level factors, such as individual body burdens of

lead. The local level refers to the immediate context that surrounds the individual, such as the concentrations of lead in the surrounding soil. The macro level refers to both the larger geospatial region that encapsulates the local level (e.g., states) and/or the broader social context (e.g., political climate and laws regarding lead-based paint in housing).

Such a conceptual framework approach underlies multilevel techniques that allow for the consideration of numerous levels *simultaneously*; that is, factors that affect health are simultaneously considered as operating at the level of the individual and the level of contexts [9]. A simple two-level study of individuals (micro level) within neighborhoods (local level) would allow us to examine whether the observed environmental health disparities were due to characteristics of individuals or characteristics of the neighborhood (context).

Figure 1: Multilevel Conceptual Framework



For example, the hypothesis that certain populations may disproportionately bear the burden of environmental contaminants prompts an important question of contextual versus individual effects. A finding that minority populations have higher lead levels is often interpreted to mean that housing quality and neighborhood deterioration within minority

neighborhoods are the accountable causes. This speaks to a contextual influence. That is, characteristics of the local environment are believed to be directly influencing individual lead levels. However, an alternative explanation is that minority populations have higher lead levels due to socioeconomic factors, health vulnerabilities, or behaviors, and that the observed associations result from the composition of individuals in those areas. From the latter perspective, lead levels would be higher among minority communities regardless of where they lived. Note that the environmental exposure need not be due to the individual or context but most likely is of the individual and context. [Note from Brad: This doesn't make sense to me] For example, the socioeconomic characteristics of place have been shown to be independently associated with individual health after accounting for individual-level socioeconomic characteristics [10]. A third level (macro level) can be incorporated into the above to additionally examine the impact of state-level policies on neighborhood characteristics and individual lead levels. Thus, multilevel approaches can improve our understanding of the complex relationships between individual and contextual influences on health [10].

In this paper, we discuss the utility of multilevel techniques to examine physical and social environmental and individual-level factors to appropriately quantify and improve our understanding of environmental health disparities. The next section of the paper discusses the reasoning and the methodological approach behind multilevel modeling, followed by considerations in choosing appropriate levels and data approaches that facilitate multilevel analyses. Our approach will be more heuristic than mathematical in order to reach a broader audience. More technical discussions can be found in numerous sources [11], [12-14], [15, 16].

MULTILEVEL MODELS

Why Multilevel Models?

Multilevel models are an extension of ordinary multiple regression that explore individual and contextual parts of variation in exposure. Multilevel models provide a mathematical modeling approach to examine between-place and between-people variability [17]. These variations are modeled by recognizing that individuals within groups, groups within local contexts, and local contexts within macro contexts may share similar characteristics.

Therefore, multilevel techniques explicitly model correlated data where the assumption of independence between observations is violated and conventional OLS techniques are not appropriate. This is in contrast to procedures that attempt to correct for the correlated structure of the data, such as those used in SUDAAN. The most commonly used multilevel statistical packages are MLwiN, HLM, and Mixed procedures within SAS. Other multilevel software include, aML, EGRET, GENSTAT, LIMDEP, LISREL, MIXREG, Mplus, R, S-Plus, SPSS, STATA, SYSTAT, and WINBUGS.

This section has been adapted from the writings of Subramanian, Duncan, and Jones [9, 18-20], who have provided leadership in applying multilevel models to the field of social epidemiology. This section is written for conceptual understanding and, therefore, mathematical and statistical details are minimized (for more technical discussions see [11], [12, 13], [14-16]).

We use the word “context” as a generic description of places or areas that may be administratively defined (as in census tracts, counties, and states) or socially defined, as in neighborhoods and communities. The notion of context can also be extended to temporal

contexts, as in different time periods. Throughout this section we use a worked example of individual lead exposure to describe how the multilevel modeling approach can potentially contribute to environmental health research by providing a mathematical modeling approach for:

- Examining a single environmental exposure that may occur through multiple media operating at different levels simultaneously and interacting at different levels.
- Examining multiple exposures operating at different levels simultaneously, potentially accumulating over time, and interacting with each other.
- Examining exposures differentially affecting subgroups of the population and/or geographic areas, and/or producing synergistic outcomes.
- Examining the fundamental role of social and economic factors and the need to account for all levels through which these mechanisms influence individual exposures, either directly or through their effect on local environments.

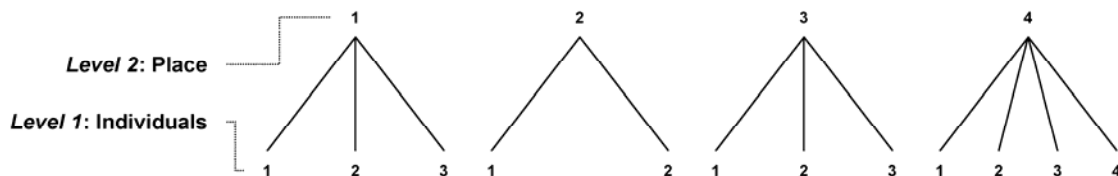
Composition and Context (Partitioning the Variance)

Research Question: Are there contextual differences in individual lead levels between contexts, after taking into account individual characteristics [9]?

Multilevel models provide the advantage of identifying and differentiating sources of variation at multiple levels, thus assigning variability to the appropriate level. It is well established that environmental exposures vary by place [21]. What is unclear is whether this variation is attributable to the composition of individuals living in that area or if these variations are independent place variations. Therefore, in starting to examine individuals within contexts, we need to identify which effects are compositional (i.e., due to the characteristics of individuals) and which are contextual (i.e., due to the characteristics of

places). In its simplest form (see Figure 2), a two-level multilevel model of individuals within places would allow us to differentiate between contextual (geographic/place) sources of variation, compositional (individual) sources of variation, and the variation due to the interaction between composition and context. Environmental studies are typically conducted at a single level, either at the aggregate/ecologic level or the micro/individual level. These studies have been critiqued due to the incorrect inference of the study results. The former relates to the well-known aggregate/ecologic fallacy when analyses are only done at the aggregate level and inference is made to the individual; the individualistic/atomistic fallacy occurs when analyses are done only at the individual level and inference is made to the group [22]. But multilevel level models are not just about avoiding fallacies; they also can provide insight into the complex processes that influence health.

Figure 2: Two-Level Structure



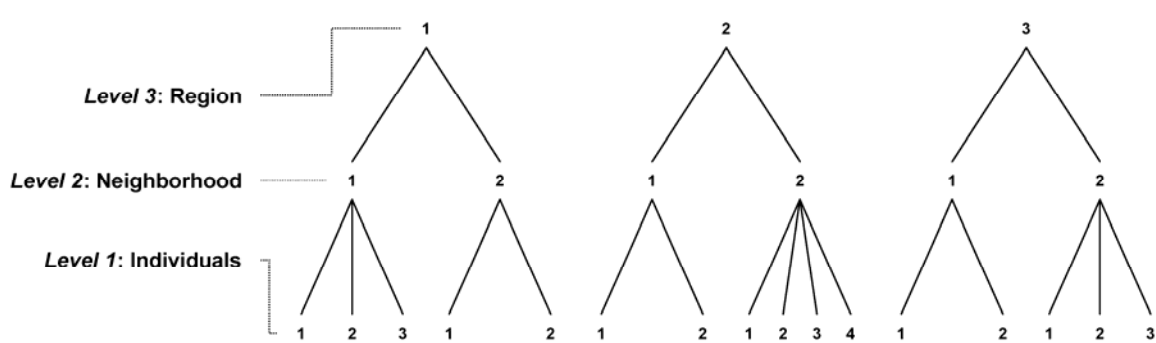
Multiple Contexts at Different Levels

Research Question: What levels are important for the study of lead exposure, and what is the relative importance of the different levels [9]?

Since multilevel models are not restricted to just two levels and potentially can be expanded to n levels, these models allow us to explore the importance of all relevant levels. For example, when the same exposure is measured at multiple levels these models allow us to evaluate the relative importance of the exposure at each level. Multilevel models summarize

the variability between higher-level units, such as the variability between neighborhoods within counties [23]. One may explore the variability of lead exposure at the neighborhood (e.g., census tract), county, and state level to ascertain which level is most important in contributing to the observed variation in individual lead levels (see Figure 3). Such an approach may also be important when examining the role of environmental policy and its implementation at the state, county, and local level, since states vary in their regulation and resource allocation for hazards such as lead exposure [24].

Figure 3: Three-Level Structure



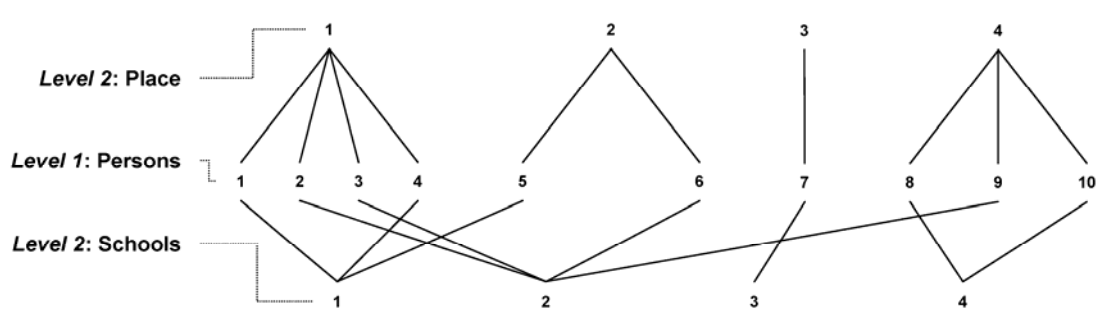
Multiple Contexts at the Same Level

Research Question: What is the relevant contribution of neighborhood and school levels, that may not be nested within one another but overlap, to lead exposure [9]?

An important feature of multilevel models is that the data need not be hierarchical. That is, contexts do not need to be neatly nested within each other. This is important as exposures commonly occur in contexts that are not hierarchical, but different contexts may occur at the same level. For example, children may be exposed to lead in the neighborhood but also within the school environment, and neighborhoods may not be nested within school districts.

In this situation we find that a number of different contexts may overlap at the same level. Such contexts are referred to as cross-classified structures. Figure 4 illustrates this concept, with individuals at level 1 nested within both neighborhoods and schools at level 2. This is important to identify contexts that may be having a confounding effect. For example, we may find that the observed variation between neighborhoods is actually variation between schools [18].

Figure 4: Multiple Contexts at the Same Level



Source: [18]

Interaction Effects

An important application of multilevel models is that it allows us to examine how variables measured at one level affect associations found at another [11]. Multilevel models allow us to examine individual and contextual interactions as well as interactions between different levels of context. These interactions are of great concern for cumulative risk studies, as in determining whether the effects of two or more exposures are merely additive or synergistic. For example, will (local-level) air pollution potentiate the effects of (individual-level) lead exposure on childhood learning disabilities? That is, among individuals with the same lead

levels is the dose response of lead amplified among those living in high-smog neighborhoods compared with those living in low-smog neighborhoods?

Individual Contextual Interaction

Research Question: What is the average association between individual lead exposure and neighborhood quality, and does this association differ for different individuals based on their poverty profile, after accounting for individual characteristics and the neighborhoods in which the individuals live [9]? That is, are poor persons living in low quality neighborhoods at higher risk of lead exposure than poor persons living in high quality neighborhoods?

Including an interaction between the individual and the context provides information about the differential effect of context across individual groups; i.e., the characteristics of individuals and of places interact to produce different effects on individual blood lead levels. Extending our current example, we may now introduce neighborhood quality and examine its association with individual poverty. We may observe that poor individuals may experience different levels of lead exposure depending on the quality of the neighborhood in which they live.

Contextual and Contextual Interaction

Research Question: What is the average association between resource allocation at the city level and neighborhood quality in relation to lead exposure, and does this association differ for neighborhoods based on their quality profile, after accounting for individual characteristics and the characteristics of the neighborhoods in which the individuals live?

Interactions may also be examined between different levels of context. Extending our example to include the level of cities with neighborhoods nested within cities, we can examine the effect of a city level variable on different types of neighborhoods. For example, including a measure of resource allocation for preventing lead exposure at the city level, we may find that for a given amount of spending, we get better results for high quality neighborhoods compared to low quality neighborhoods. Such observations are important for policy development and resource allocation for preventing environmental exposures.

Modeling Contexts and Individuals Over Time

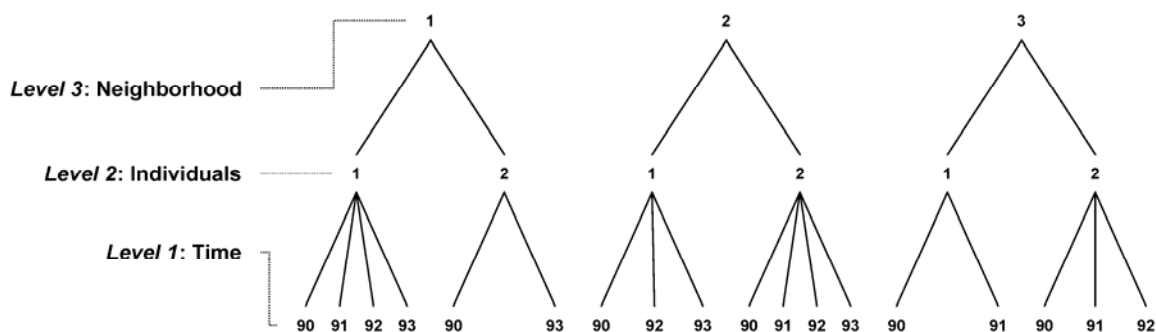
Multilevel models allow us to examine changes over time, an important aspect in monitoring environmental health disparities. As contexts change over time, so do the exposure levels in individuals. There are two possible situations depending on whether individuals are repeatedly measured or the context is repeatedly measured. While the use of multilevel analysis of individuals nested within contexts is fairly intuitive, the repeated measurement of contexts is not.

Repeated Measures of Individuals

Research Question: While individual lead exposures may have declined over time, have neighborhood contextual disparities declined or increased, and for which population groups have the contextual disparities declined or increased [9]?

If individuals are measured repeatedly over time, as in a panel design, then their measurements can be described as being nested within each individual. For example, we may monitor lead levels of a group of individuals over time, so we would have each lead level for each time period at level 1 nested within individuals at level 2 nested within neighborhoods at level 3, as seen in Figure 5. The advantage of the multilevel design over conventional repeated measures analyses is that the number of measurements per individual, as well as the spacing between measurements, may vary. In this case we are examining individual change within a contextual setting [18].

Figure 5: Repeated Measures of Individuals



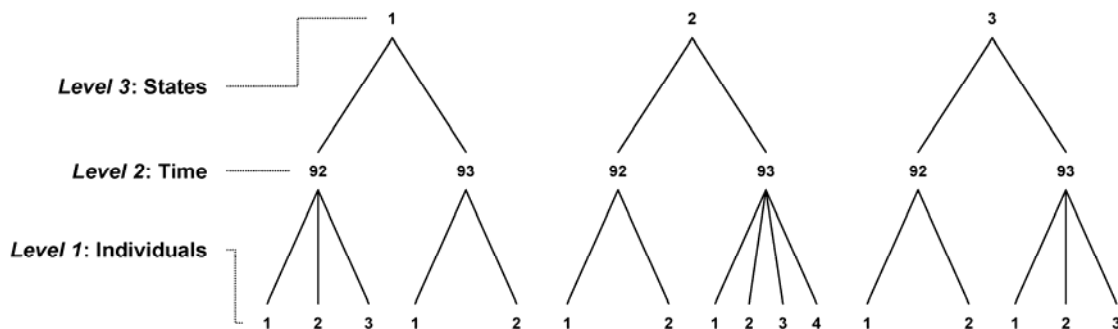
Source: [18]

Repeated Measures of Contexts

Research Question: Which types of individuals and which types of places have changed over time with respect to lead exposure [9, 25]?

If repeated cross-sectional surveys are conducted within a certain context, they can be regarded as repeatedly measuring contexts over time. This design can be described as individuals nested within time, nested within contexts. For example, statewide surveys that are conducted annually will produce individuals nested within time nested within states as shown in Figure 6. In this case, we can examine trends within states while controlling for individual characteristics.

Figure 6: Repeated Measures of Contexts



Source:[18]

Finally, multilevel models require that the spatial autocorrelation (the similarity between individuals for a given variable as a function of spatial distance) be accounted for, e.g., as in point sources of pollution. The combination of spatial models with multilevel models is relatively new, but published studies that incorporate structures that describe spatial adjacency are already available [26] [27].

Types of Contextual Variables Used in Multilevel Models

While we have discussed the use of contextual variables, it is important to note that contextual variables can be measured and interpreted in different ways; this section introduces some commonly defined contextual variables. These variables are typically used to describe the characteristics of a group or context.

Derived variables are contextual variables that are summarized from the characteristics of individuals in that context, such as median neighborhood income or the percentage of high school graduates in a neighborhood. Some derived variables have no individual level equivalent, such as inequalities in the income distribution in an area, while others, such as average neighborhood income, do. While derived variables may be summarized from individual characteristics, their effect may be independent in that, conceptually, they may be measuring a characteristic of the context. The term “derived variable” is used synonymously with analytic and aggregate variables [22, 28].

Integral variables, in contrast to derived variables, do not have an individual equivalent. Zoning policies, racial segregation, and population density are examples of integral variables. Thus, integral variables often describe group properties that are distinct from properties of the individuals comprising these groups [22, 28].

Environmental variables within the context of multilevel models have been described as measures of physical and chemical exposures. Environmental variables are generally not aggregated from individual-level variables but do have individual-level equivalents. Such variables are typically used as proxies for individual-level variables that may be difficult to measure at the individual level [22, 28]. A common example of an environmental variable used as a proxy for an individual variable is the use of the ambient outdoor concentration of an air pollutant as a proxy for the personal exposure concentration of the local residents.

Because (a) most people spend more than 80% of their time indoors or in vehicles, where pollutant concentrations can be significantly different from those outdoors, and (b) many people spend a substantial amount of time at locations other than their residence, the personal exposure concentration can be quite different from the ambient outdoor concentration in the vicinity of the residence. As explained in Appendix C, modeling tools and databases are available to estimate personal exposure concentrations. Multilevel methods advocate measuring the exposure at both the level of the residence and the level of the individual.

This section has provided a brief overview of multilevel models. While multilevel models can provide an important tool to improve our understanding of environmental exposure and its relationship to socioeconomic position, the primary research needs to be theoretically justifiable, and model complexity needs to be balanced with functional applications.

IDENTIFYING THE RELEVANT LEVELS

This section addresses the questions and principles that guide the choice of levels or geographic units in multilevel studies. These include the research question being addressed, the theoretical pathways linking the micro-local and macro levels, health outcomes and exposures under consideration, data availability, and the administrative or intervention application of the research. Social theories underlying choice of levels are discussed in detail in Appendix A. Quantifying the importance of different levels using the multilevel approach can provide insights into the theoretical pathways linking the micro-local-macro processes. Direct policies, synergistic policy effects, sensitivity periods, and mediating mechanisms will also necessarily influence the definition of the relevant levels [29].

Although we have tried to use conventions such as “individuals” to refer to the micro level, a multilevel approach is not necessarily restricted to the micro level defined as individuals or persons; the micro level can refer to industrial facilities or other theoretically defined units. The classic example of environmental justice examines the placement of industrial facilities in communities, with the central argument that race is a key determinant of facility location [30]. If we consider facilities as the micro level, we can consider whether racial composition at the local level influences the placement of facilities, net of other economic characteristics. Further, we could consider the influence of municipal or state policies as determinants of local racial composition and facility placement.

We briefly discuss census geography, as it is the most commonly used source to identify contextual levels. Following this discussion, issues that guide the choice of levels will be addressed.

Level Specification

Table 1 provides examples of geographic levels and micro levels that have been used and/or potentially can be considered in multilevel studies [31, 32]. The most commonly used contextual levels for multilevel analyses are based on census geography and are italicized in Table 1. Census geography is built hierarchically; e.g., from the block level representing 85 individuals, up to the block group with an average population of 1500, to census tracts that represents about 4000 people, to county subdivisions, to counties and then states or from school districts to states (a more detailed description of census geographic units is provided in Appendix B). Block groups and census tracts have been used to approximate neighborhoods, as their boundaries are defined to create population groups that are homogenous with regard to social and economic characteristics [31, 32]. Although Zip codes have commonly been used in health research, their use is not recommended. While census tracts and block groups are delineated to be homogenous units, zip codes are defined by the US Postal Service (USPS) for efficient mail delivery and can range in size from a single building to large areas that cross state boundaries. The use of Zip codes is further complicated by extensive modifications in the past 10 years, the creation of new zip codes, and the deletion of existing zip codes. To overcome these discrepancies, the 2000 US Census defined Zip Code Tabulation Areas (ZCTA) mapped to census blocks to replace zip codes. The current 5-digit ZCTA area may no longer correspond to the USPS 5-digit zip code area and no relational files between zip codes and ZCTA will be released by census [33].

The Traffic Analysis Zone (TAZ), the primary unit to describe traffic data, may be an important unit for the study of environmental health disparities. TAZ consists of one or more census blocks, block groups, or census tracts and is defined by states and/or local

transportation officials. Macro- level units such as states and counties are important in policy evaluation efforts.

Table 1: Examples of Units at Different Levels

Macro levels	Local levels	Micro levels
Countries	Change in place (time)	Individual growth (time)
EPA regions	Communities	Cellular matrices
<i>States</i>	Neighborhoods	Blood markers
<i>MSAs</i>	Work places	DNA
<i>Counties</i>	Zoning districts	Observations
<i>Cities</i>	<i>Voting districts</i>	Individuals
<i>State economic areas</i>	<i>Congressional districts</i>	Families
<i>Labor market areas</i>	<i>School districts</i>	Homes
<i>State legislative districts</i>	<i>Traffic analysis zones</i>	Schools
	<i>Census tracts</i>	Industrial facilities

Zoning districts may also be an important unit for the study of environmental health disparities, as zoning is a commonly used planning tool that can influence the location of environmental hazards. Areas zoned for industrial use have been shown to be concentrated in poor and minority neighborhoods, resulting in higher environmental burdens for these neighborhoods. Zoning districts and their relationship to neighborhoods and/or census tracts should be considered important levels in environmental health research [34, 35].

Populations Not Defined by Conventional Levels

Conventional geographic units (e.g., census tracts) are likely inadequate indicators of the “contexts” for homeless and migrant populations (e.g., migrant workers). However, some creativity in operationalizing these conventional units can be informative, e.g., Culhane et. al. (1996) [36] present an interesting methodology using census tracts for studying homeless populations. Further, emerging immigrant populations (e.g., Arab Americans) may

form and change their communities between census periods, creating a potential disjunction between the local- and macro-level measurements. This problem may potentially be addressed by The American Community Survey, which will provide comparable data to that available on the US Census on an annual basis for all states, cities, counties, and metropolitan areas, and every five years for smaller geographic units. Also, most of the research using census-derived data focuses on metropolitan areas, leaving relatively unexplored the applicability of these indicators for use with rural populations. Clearly, however, multilevel studies have yet to address the relevant levels or contexts and contextual measures for these population groups.

Issues to Be Considered in Level Specification

The Research Question

Conceptually, the research agenda needs to be broadened to include the macro-local and micro-level processes through which socioeconomic health differentials are structured. It is well recognized that disadvantage is locally concentrated. The question to address is: What are the higher-level precursor processes that result in the observed geographic clustering of economic disadvantage, social disadvantage, vulnerability, and environmental exposures [32], and how do micro-level characteristics of individuals tie in with the above? Such inquiries provide a holistic approach to studying socioeconomic disparities and from a levels perspective are more likely to ensure that all relevant levels are considered.

Health Outcome

The health outcome(s), if any, under study influences the level(s) at which the analyses are done. The causal links between health outcomes and physical and social environments are

typically complex and multifactorial, requiring multiple levels of inquiry. Certain outcomes may be more variable at the micro or local levels, while others may be more dependent on the macro level. The level of dependence of the health outcome may also be different for racial/ethnic groups. For example, census tract level variability in mortality was found to be six times greater for blacks than for whites [37]. Existing health conditions by racial/ethnic group, given higher rates of certain conditions among minority populations, may further influence the choice of level.

Social Definitions vs. Administrative Definitions vs. Intervention Levels

How an area is defined needs also to be considered in the context of its social and administrative function. Residents' perceptions of a neighborhood may not coincide with census tracts. Thus, while administratively convenient, census tracts may fail to capture, or even misclassify, important aspects of the social, physical, and built environments. Further, some potential exists that ethnic and socioeconomic groups may vary in their perceptions of neighborhood boundaries [3]. We need to also consider how these boundaries are defined by governmental or non-governmental agencies in defining neighborhoods or communities for when intervention targeting is planned. For example, the state may be an important level for outcomes that are directly influenced by state policies, while city planning departments typically define neighborhood boundaries along census tract boundaries.

Modifiable Area Unit Problem (MAUP)

With the increase in multilevel studies, the issue of the modifiable area unit problem (MAUP) has received much attention. This relates to the definition of the level(s) (area unit) into which individuals or areas are aggregated. This is an important consideration since changes in the size and number of levels (areas) can influence the observed associations

significantly, reversing the direction of the observed relationships in the extreme case [38, 39].

Exposure Period

While the choice of levels is an important consideration in multilevel models, the effect of the social and physical environment on individuals is dependent on the exposure period within the local environment. The underlying assumption in multilevel models is that individuals in a given area share common exposures and experiences. Presumably, the local environment is a less relevant level for individuals whose work, school, or social activities are outside of the local area. Population exposure modeling using population activity data does account for variability in exposure periods for different age groups, which will be discussed later.

However, the spatial pattern of daily life in relation to the definition of the local environment remains an outstanding area of research [40]. Further, some research has suggested that minorities have a smaller geographic range, which may lead to greater variability in exposure [41].

Environmental Exposures

Characteristics, measurement, multiple sources, relevant policies, and the research goal for environmental exposure influence the choice of the analytic levels. Outdoor air pollutants and ambient and drinking water quality are used as examples to provide an in-depth description of how these issues influence the choice of levels (see Appendices C and D).

For example, estimation procedures directly influence the level at which data are available, with air quality monitoring data available at more aggregate levels such as counties and dispersion modeled data potentially available at the level of individual address. In the case of water quality, using water monitoring or modeling data with demographic data at the

census tract level can be difficult. For example, in a study of problems associated with collecting drinking water quality data for community studies, the task of evaluating water quality for each census tract was complicated by the fact that single census tracts were served by more than one system [42]. This is a classic problem that the multilevel modeling approach (see “Multiple Contexts at the Same Level” section) can potentially address.

DATA LINKAGES FOR MULTILEVEL MODELS

Data requirements for multilevel models require, at a minimum, that observations have identifiers that differentiate the contextual setting(s) of each observation. Thus multilevel data sources are essential for multilevel analyses. This section describes the ways in which different data sources can be combined to produce suitable data for multilevel analyses. Census data is the most commonly used source for characterizing and defining contexts; therefore most national surveys are now being routinely geocoded and linked to census data to facilitate multilevel analyses. In addition, because of how data are collected, as for environmental exposures, or because of the sampling strategy of national surveys, geographic identifiers are readily available on some datasets. Table 2 provides data source examples where the contextual identifiers are available with monitoring, importantly the appropriateness of the level that data is available at should be evaluated as discussed earlier (see “Identifying the Relevant Levels”). Table 2 also provides examples where survey respondent data have been linked to contextual levels (e.g., NHIS and NHANES).

Geocoding

Geocoding individual records is an important step towards creating multilevel data sources. Geocoding refers to the process by which individual addresses are assigned a corresponding set of latitude and longitude coordinates. These coordinates can then be coded to any geographic unit by determining within which geographic unit the specific coordinates are based. For example, these coordinates can be mapped to their respective census blocks, census tracts, and EPA areas.

Because geocoding requires a physical address, this approach is especially problematic for undeclared immigrants and homeless and mobile populations and requires further methodological development to appropriately contextualize these populations.

Table 2: Sample Data Sources with Geographic Identifiers

Data	Indicators	Available Levels
<i>Environmental exposures</i>		
Aerometric Information Retrieval System (AIRS)	Common air pollutants	County, MSA, state
National Air Toxics Assessment (NATA)	Hazardous air pollutants	Census tract, County, MSA, state
Toxic Release Inventory data (TRI)	Toxic chemical releases from industrial facilities	Individual facilities, county, state
Safe Drinking Water Information System (SDWIS)	(a) Drinking water contaminants (b) Violations of monitoring	Water system, county, MSA, state
Superfund NPL Assessment Program (SNAP)	Residence in relation to site	Site locations, county, MSA, state
<i>Body burdens</i>		
National Health and Nutrition Examination Survey (NHANES)	(a) Mercury in women of child bearing age (b) Lead and continine in children's bloods	Block group, tract, county, MSA, state
Health indicators		
National Health Interview Survey (NHIS)	(a) Asthma prevalence (b) Preexisting health conditions (c) Asthma emergency room visits	Block group, tract, county, MSA, state
Surveillance, Epidemiology	Cancer by type	Block group, tract, (CA only)
End Results Program (SEERS)	Cancer by type	Specific states and MSAs

Contextual identifiers pose a serious confidentiality problem. Because block groups are relatively small areas—when rare health outcomes are considered and cross tabulated with other covariates—the potential to identify the respondent within the block group is very high. Therefore, when the decision is made to geocode, restrictions need to be in place to protect the confidentiality of individuals. General approaches to reduce confidentiality breaches include anonymized identifications, cell suppression, aggregation, top coding, and reducing the detail for categorical variables [43]. To facilitate the analysis of contextual identifier

linked national health surveys, the National Center of Health Statistics has created Research Data Centers.

The issue of geocode accuracy is not adequately addressed in most datasets or studies. Very large discrepancies in accuracy have been noted between commercial firms, so researchers are encouraged to report and evaluate the accuracy of the geocoding methods [44].

Examples of Data Linkages

This section provides some examples of how different data sources can be linked to create multilevel data structures including census data linked to survey data, census and tax assessor data linked to state health department data, and national health survey data linked to environmental exposure data.

Contextual Variables Linked to Geocoded Data

The geocoding of individual records allows for these records to be linked with contextual variables. Contextual variables from census data may be linked with individual records (e.g., of the NHIS and NHANES) at the tract, county, city, MSA, and state levels. Contextual variables from The Area Resource File on health services, economic, and environmental characteristics can also be linked to these surveys at the county level. We recommend that such datasets at minimum routinely append information on neighborhood poverty and racial composition, even if geographic identifiers are not released.

State Blood Lead Screening Data Linked to Tax Assessor Data

Miranda and colleagues (2002) [45] propose a promising approach to geocoding North Carolina blood lead screening data to the tax parcel. The tax parcel is the land entitlement

unit, and for residential units the tax parcel describes the housing structure (single or multifamily) and surrounding yard. Tax assessor data can potentially provide variables such as year of construction, tax value, date remodeled, and renter/owner occupied, quantifying the value and overall maintenance of the housing structure. Since age of housing can vary even within a census block, geocoding children to their actual residential unit is advantageous. Note that the quality of tax assessor data should be examined.

State Blood Lead Screening Data Linked to Census Data

Krieger and colleagues (2003) [46] have proposed a geocoding strategy and the use of area-based measures based on census data for monitoring socioeconomic position in the absence of socioeconomic data for state blood lead screening data. The authors advise that the best suited census tract and census block group socioeconomic measures are those 1) most sensitive to capturing economic deprivations, 2) meaningful for comparison across regions and over time, and 3) easily interpretable with categorical cut points. The study recommends the use of census tracts, census blocks groups, and poverty-related measures based on the strongest socioeconomic gradients observed. This approach has been validated using multiple geographic levels (block group, census tract, and zip code), area-based measures (occupational class, wealth, poverty, income, education, crowding, and composite indices), and health outcomes (low birth weight, infectious disease and injury, cancer incidence, and all-cause and cause-specific mortality). The underlying assumption of this approach is that these area-based measures capture a mixture of individual-level and area-level socioeconomic effects. Given the data limitation of this approach, the effects of composition, context, and the interaction between composition and context cannot be differentiated.

NHANES Linked to Aerometric Information Retrieval System (AIRS)

Schwartz (2001) [47] presents a multilevel analytic study using NHANES linked to AIRS data. This approach was adopted to examine the association between blood markers (fibrinogen level, platelet counts, and white blood cell count) and air pollution (PM₁₀, sulfur dioxide, and nitrogen dioxide). As this study demonstrates, associations can be made from the macro level through to the intra-individual level.

The development of appropriate multilevel data sources is an important step towards the successful application of multilevel techniques to quantify and monitor socioeconomic and racial/ethnic disparities in environmental health.

CONCLUSION

Eliminating, rather than merely reducing, racial/ethnic and socioeconomic disparities in health is a major U.S. health policy objective. This objective, coupled with extreme residential segregation by race/ethnicity and SEP experienced by the U.S. population, calls for the need to incorporate innovative approaches to examining risks occurring at multiple levels and over time.

Multilevel models can potentially contribute to environmental health disparities research and monitoring by providing an analytic approach for:

- Informing environmental policies and examining the impact of existing policies on local contexts and individual exposures.
- Examining a single environmental exposure that may occur through multiple media operating at different levels simultaneously and interacting at different levels.
- Examining multiple exposures operating at different levels simultaneously, potentially accumulating over time, and interacting with each other.
- Examining exposures differentially affecting subgroups of the population and/or geographic areas, and/or producing synergistic outcomes.
- Examining the fundamental role of social and economic factors and the need to account for all levels through which these mechanisms influence individual exposures, either directly or through their effect on local environments.
- Developing theoretical models to explain environmental health disparities or to generate hypotheses.

Although numerous challenges in multilevel research remain, we call attention to the emerging data and approaches for assessing the convergence of social, economic,

racial/ethnic, and environmental factors operating at multiple levels simultaneously in generating and sustaining environmental health disparities.

We can offer several suggestions to facilitate the integration of multilevel techniques in the study of environmental health disparities:

- a) Choice of level(s) will partially depend upon existing data. Gaps in data that are needed to develop and apply multilevel models can be filled through collection of new data and converting existing sources into accessible and standardized formats.
- b) Making data sets better able to be cross-referenced with uniform identification codes.
- c) Development of a standard multilevel census data set with SEP and other data, which can be merged with population-based data sources such as NHANES, and then have geographic IDs anonymized to maintain privacy and confidentiality.
- d) Datasets at minimum routinely append information on neighborhood poverty and racial composition, even if geographic identifiers are not released.
- e) Increasing the capacity for multilevel techniques through training and interdisciplinary research teams.

BIBLIOGRAPHY

1. White House Office of President, *Federal actions to address environmental justice in minority populations and low-income populations*, in *Executive Order 12898*. 1994.
2. McMichael, A.J., *Prisoners of the proximate: loosening the constraints on epidemiology in an age of change*. *Am J Epidemiol*, 1999. **149**(10): p. 887-97.
3. Williams, B.L., Y. Florez, and S. Pettygrove, *Inter- and intra-ethnic variation in water intake, contact, and source estimates among Tucson residents: Implications for exposure analysis*. *J Expo Anal Environ Epidemiol*, 2001. **11**(6): p. 510-21.
4. Evans, G.W. and E. Kantrowitz, *Socioeconomic status and health: the potential role of environmental risk exposure*. *Annu Rev Public Health*, 2002. **23**: p. 303-31.
5. U.S. Environmental Protection Agency, *Framework for cumulative risk assessment*. 2003, Office of research and development: Washington, DC.
6. Schulz, A. and M.E. Northridge, *Social determinants of health: implications for environmental health promotion*. *Health Educ Behav*, 2004. **31**(4): p. 455-71.
7. Morello-Frosch, R., et al., *Environmental justice and regional inequality in southern California: implications for future research*. *Environ Health Perspect*, 2002. **110 Suppl 2**: p. 149-54.
8. Gee, G.C. and Payne-Sturges, *Environmental health disparities: a framework integrating psychosocial and environmental concepts*. *Environ Health Perspect*, 2004. **112**(17): p. 1645-53.
9. Subramanian, S.V., K. Jones, and C. Duncan, *Multilevel methods for public health research*, in *Neighborhoods and health*, I. Kawachi and L. Berkman, Editors. 2003, Oxford University Press: New York, NY. p. 65-111.
10. Macintyre, S., A. Ellaway, and S. Cummins, *Place effects on health: how can we conceptualise, operationalise and measure them?* *Soc Sci Med*, 2002. **55**(1): p. 125-39.
11. Bryk, A.S. and S.W. Raudenbush, *Hierarchical Linear Models: Applications and Data Analysis Methods*. 1992: Newbury Park, England, Sage Publications.
12. De Leeuw, J. and I. Kreft, *Software for Multilevel Analysis*. 1999: p. 1-22.
13. Singer, J.D., *Using SAS PROC MIXED to Fit Multilevel Models, Hierarchical Models, and Individual Growth Models*. *The Journal of Educational and Behavioral Statistics*, 1997(December).
14. Goldstein, H., *Multilevel Statistical Models*. 2nd ed. 1995: Arnold, London.
15. Bingenheimer, J.B. and S.W. Raudenbush, *Statistical and substantive inferences in public health: issues in the application of multilevel models*. *Annu Rev Public Health*, 2004. **25**: p. 53-77.
16. Hox, J. and I. Kreft, *Multilevel Analysis Methods*. *Sociological Methods and Research*, 1994(22): p. 283-299.
17. Duncan, C. and K. Jones, *Using multilevel models to model heterogeneity: potentials and pitfalls*. *Geographical Analysis*, 2000. **32**(4).
18. Duncan, C., K. Jones, and G. Moon, *Health-related behavior in context: a multilevel modeling approach*. *Soc Sci Med*, 1996. **42**(6): p. 817-30.
19. Duncan, C., K. Jones, and G. Moon, *Context, composition and heterogeneity: using multilevel models in health research*. *Soc Sci Med*, 1998. **46**(1): p. 97-117.
20. Jones, K., *Specifying and Estimating Multilevel Models for Geographic Research*. *Transactions of the Institute of British Geographers, New Series*, 1991(16): p. 148-160.
21. Lopez, R., *Segregation and black/white differences in exposure to air toxics in 1990*. *Environ Health Perspect*, 2002. **110 Suppl 2**: p. 289-95.
22. Diez-Roux, A.V., *Research Forum: Bringing Context Back into Epidemiology: Variables and Fallacies in Multilevel Analysis*. *American Journal of Public Health*, 1998. **88**(2): p. 216-222.
23. Rice, N. and A. Jones, *Multilevel Models and Health Economics*. *Health Economics*, 1997. **6**(6)(November-December): p. 561-75.
24. Brown, M.J., et al., *The effectiveness of housing policies in reducing children's lead exposure*. *Am J Public Health*, 2001. **91**(4): p. 621-4.
25. Subramanian, S.V., *The relevance of multilevel statistical methods for identifying causal neighborhood effects*. *Soc Sci Med*, 2004. **58**(10): p. 1961-7.
26. Burnett, R., et al., *The spatial association between community air pollution and mortality: a new method of analyzing correlated geographic cohort data*. *Environ Health Perspect*, 2001. **109 Suppl 3**: p. 375-80.

27. Langford, I.H., G. Bentham, and A.L. McDonald, *Multilevel Modeling of Geographically Aggregated Health Data: A Case Study on Malignant Melanoma Mortality and UV Exposure in the European Community*. *Statistics in Medicine*, 1998(17): p. 41-57.
28. Diez Roux, A.V., *A glossary for multilevel analysis*. *J Epidemiol Community Health*, 2002. **56**(8): p. 588-94.
29. Diez Roux, A.V., *Investigating neighborhood and area effects on health*. *Am J Public Health*, 2001. **91**(11): p. 1783-9.
30. Perlin, S.A., D. Wong, and K. Sexton, *Residential proximity to industrial sources of air pollution: interrelationships among race, poverty, and age*. *J Air Waste Manag Assoc*, 2001. **51**(3): p. 406-21.
31. Diez Roux, A.V., *Estimating neighborhood health effects: the challenges of causal inference in a complex world*. *Soc Sci Med*, 2004. **58**(10): p. 1953-60.
32. Brooks-Gunn, J., G.J. Duncan, and L.J. Aber, *Neighborhood Poverty: Context and Consequences for Children.*, ed. R.S. Foundation. Vol. 1. 1997, New York. 334.
33. Krieger, N., et al., *Zip code caveat: bias due to spatiotemporal mismatches between zip codes and US census-defined geographic areas--the Public Health Disparities Geocoding Project*. *Am J Public Health*, 2002. **92**(7): p. 1100-2.
34. Maantay, J., *Zoning, equity, and public health*. *Am J Public Health*, 2001. **91**(7): p. 1033-41.
35. Maantay, J., *Zoning law, health, and environmental justice: what's the connection?* *J Law Med Ethics*, 2002. **30**(4): p. 572-93.
36. Culhane, D.P., C. Lee, and S.M. Wachter, *Where the homeless come from: A study of the prior address distribution of homeless families admitted to homeless shelters in New York City and Philadelphia, in Housing Policy Debate*. 1996, FannieMae Foundation.
37. Subramanian, S.V., et al., *Racial disparities in context: a multilevel analysis of neighborhood variations in poverty and excess mortality among black populations in Massachusetts*. *Am J Public Health*, 2005. **95**(2): p. 260-5.
38. Heywood, *Introduction to geographical information systems*. 1998, New York: Addison Wesley Longman.
39. Armhein, C.G., *Searching for the elusive aggregation effect: Evidence from statistical simulations*. *Environment & Planning A*, 1994. **27**(1): p. 105-119.
40. Sastry, N., A.R. Pebley, and M. Zonta, *Neighborhood Definitions and the Spatial Dimension of Daily Life in Los Angeles*. 2002, Labor and Population Program.
41. Massey, D., *Residential segregation and neighborhood conditions in U.S. metropolitan areas*, in *America becoming: racial trends and their consequences*, N.J. Smelser, W.J. Wilson, and F. Mitchell, Editors. 2001, National Academy Press: Washington, D.C.
42. Whorton, M.D., et al., *Problems associated with collecting drinking water quality data for community studies: a case example, Fresno County, California*. *Am J Public Health*, 1988. **78**(1): p. 47-51.
43. Cox, L.H., *Protecting confidentiality in small population health and environmental statistics*. *Stat Med*, 1996. **15**(17-18): p. 1895-905.
44. Krieger, N., et al., *On the wrong side of the tracts? Evaluating the accuracy of geocoding in public health research*. *Am J Public Health*, 2001. **91**(7): p. 1114-6.
45. Miranda, M.L., D.C. Dolinoy, and M.A. Overstreet, *Mapping for prevention: GIS models for directing childhood lead poisoning prevention programs*. *Environ Health Perspect*, 2002. **110**(9): p. 947-53.
46. Krieger, N., et al., *Choosing area based socioeconomic measures to monitor social inequalities in low birth weight and childhood lead poisoning: The Public Health Disparities Geocoding Project (US)*. *J Epidemiol Community Health*, 2003. **57**(3): p. 186-99.
47. Schwartz, J., *Air pollution and blood markers of cardiovascular risk*. *Environ Health Perspect*, 2001. **109 Suppl 3**: p. 405-9.
48. O'Neill, M.S., et al., *Health, wealth, and air pollution: advancing theory and methods*. *Environ Health Perspect*, 2003. **111**(16): p. 1861-70.
49. Krieger, N., *Theories for social epidemiology in the 21st century: an ecosocial perspective*. *Int J Epidemiol*, 2001. **30**(4): p. 668-77.
50. Cassel, J., *The contribution of the social environment to host resistance: the Fourth Wade Hampton Frost Lecture*. *Am J Epidemiol*, 1976. **104**(2): p. 107-23.
51. McEwen, B.S., *Protective and Damaging Effects of Stress Mediators*. *Seminars in Medicine of the Beth Israel Deaconess Medical Center*, 1998. **338**(3): p. 171-179.
52. Krieger, N., *Epidemiology and the Web of Causation: Has Anyone Seen the Spider?* *Social Science and Medicine*, 1994. **39**(7): p. 887-903.

53. Subra, W. *The National Environmental Justice Advisory Council's Report on Environmental Justice and Cumulative Risks and Impacts*. in *Science of Environmental Justice Conference*. 2004. Boston University, Boston, MA.
54. Dietert, R.R., et al., *Workshop to identify critical windows of exposure for children's health: immune and respiratory systems work group summary*. *Environ Health Perspect*, 2000. **108 Suppl 3**: p. 483-90.
55. Luster, M.I., J.H. Dean, and D.R. Germolec, *Consensus workshop on methods to evaluate developmental immunotoxicity*. *Environ Health Perspect*, 2003. **111**(4): p. 579-83.
56. Bhattacharya, J., et al., *Heat or eat? Cold-weather shocks and nutrition in poor American families*. *Am J Public Health*, 2003. **93**(7): p. 1149-54.
57. Galobardes, B., J.W. Lynch, and G. Davey Smith, *Childhood socioeconomic circumstances and cause-specific mortality in adulthood: systematic review and interpretation*. *Epidemiol Rev*, 2004. **26**: p. 7-21.
58. Macintyre, S., S. Maciver, and A. Sooman, *Area, class and health: should we be focusing on places or people?* *J. Soc. Pol.*, 1993. **22**(2): p. 213-234.
59. Sampson, R.J., S.W. Raudenbush, and F. Earls, *Neighborhoods and Violent Crime: A Multilevel Study of Collective Efficacy*. *Science*, 1997. **277**(August): p. 918-924.
60. Braveman, P.A., *Monitoring equity in health and healthcare: a conceptual framework*. *J Health Popul Nutr*, 2003. **21**(3): p. 181-92.
61. Robinson, J.G., *ESCAP II: Demographic Analysis Results*. 2001, Bureau of the Census: Washington, DC.
62. William, G.B., *Preliminary Estimates Show Improvement in Census 2000 Coverage, in United States Department of Commerce News*. 2001: Washington, DC.
63. US Census Bureau, *American Community Survey: What is the American Community Survey?*
64. US Department of Commerce, Economics and Statistics Administration, and US Census Bureau, *Geographic Areas Reference Manual*. 1994.
65. US Census Bureau, *Standard hierarchy of census geographic entities*.
66. Krieger, N., et al., *Painting a Truer Picture of US Socioeconomic and Racial/Ethnic Health Inequalities: The Public Health Disparities Geocoding Project*. *Am J Public Health*, 2005. **95**(2): p. 312-23.
67. EPA, *Drinking Water From Household Wells*. 2002.
68. Lee, S., et al., *Drinking Water Exposures and Perceptions among 1998-1999 FoodNet Survey Respondents*. 2002, International Conference on Emerging Infectious Diseases: Atlanta, GA.
69. Whitaker, H., et al., *Modeling exposure to disinfection by-products in drinking water for an epidemiological study of adverse birth outcomes*. *J Expo Anal Environ Epidemiol*, 2004.
70. McLaughlin, D.K., C.S. Stokes, and A. Nonoyama, *Total and Race-specific Mortality in the U.S.: Inequality Across U.S. Counties*. 2001, Department of Agricultural Economics and Rural Sociology and Population Research Institute, The Pennsylvania State University, Population Association of America Annual Meeting: Washington, DC.
71. Natural Resources Defense Council (NRDC), *Hidden Danger-Environmental Health Threats in the Latino Community*. 2004.

APPENDIX A: PLACE AND SOCIAL THEORY

Many theoretical frameworks exist for examining the roles of social and physical environments in racial/ethnic and socioeconomic health disparities. These frameworks vary in name, disciplinary origin, and emphasis, but they share some common themes.

Several theoretical frameworks for understanding environmental health disparities have recently been proposed. Gee and Payne-Sturges (2004) [8] build upon the exposure-disease paradigm in suggesting that psychosocial stress is the key component that explains the greater susceptibility of disadvantaged populations to environmental hazards. Calling it the stress-exposure disease framework, they hypothesize that residential segregation is the reason that “race” is important, incorporate a multilevel perspective, and argue that racial differences in stressors account for racial differences in vulnerability [8]. Schulz and Northridge (2004) [6] have developed a framework for understanding social and environmental inequalities in health, drawing on earlier frameworks intended to understand racial disparities in health and incorporating factors in the built environment. In this framework, macro factors (e.g., structural determinants such as distribution of wealth) influence and are influenced by local factors (e.g., land use and community investment), which then influence and are influenced by proximate factors (e.g., health behaviors and housing conditions), which ultimately influence health and well-being [6]. Morello-Frosch and colleagues (2002) [7] suggest that income inequality and social capital at the macro levels affect the ability of local communities to influence environmental and social policies and, consequently, their ability to resist environmental health stressors such as the placement of hazardous waste facilities and subsequent health effects.

SEP is a key variable in the theories and frameworks summarized above. For example, a recent paper by O'Neill and colleagues (2003) [48] illustrates how social epidemiology theory can be incorporated in understanding disparities in air pollution: “First, groups with lower SEP may have higher exposure to air pollution. Second, because lower-SEP groups already experience compromised health status... they may be more susceptible to the health effects of air pollution. Third, because of the combination of greater exposure and susceptibility, these groups are likely to suffer greater health effects from air pollution exposure.” Racial/ethnic environmental health disparities can be viewed in a similar manner.

In modern social epidemiology, prominent theoretical models include psychosocial theory, social production of health, and ecosocial theory and its related frameworks [49]. Cassel (1976), a leader in defining psychosocial theory, proposed that psychosocial factors in the social environment (e.g., isolation, disorganization, and support) influence the degree to which some persons become more or less susceptible to disease-causing agents than others [50]. In this model, the social environment affects health indirectly by influencing susceptibility through changes in the neuroendocrine function. Others have proposed that psychosocial factors can affect health directly through allostatic load, the “wear and tear” of organ systems resulting from stressors [51]. The social production of health perspective, sharing much in common with the political economy of health theory, stresses “upstream” political and economic determinants of health, including income inequality, racial discrimination, neoliberal economic policies, and deregulation of corporations [49]. In part a reaction to proponents of individual responsibility for health, the social production of health theory claims that class inequalities are the fundamental causes of health inequalities. Ecosocial theory [49, 52] and its related multilevel frameworks incorporate multiple levels of organization (biological, social, and ecological) over time and space to explain changing population patterns of health [49]. The intent of ecosocial theory is to provide a set of

guiding principles for scientific inquiry and to incorporate accountability (e.g., institutions), as does the social production of health theory.

There are several common themes in the theories and frameworks discussed above. Of note are: (1) the concept of vulnerability, (2) multiple levels/nested hierarchies, and (3) the incorporation of time. “Vulnerability” is a defining concept in the field of environmental justice and has been categorized into at least four overlapping types: susceptibility/sensitivity (e.g., vulnerable populations such as children or the elderly), differential exposure (e.g., proximity to pollution sources), differential preparedness (e.g., low income), and differential ability to recover (e.g., discrimination) [53]. Vulnerability is key to understanding both racial/ethnic and socioeconomic disparities in environmental health.

In addition, these theories and frameworks consider multiple levels and/or nested hierarchies. In psychosocial theory, emphasis is placed on at least two levels, the characteristics of individuals and of the social environment. The social production of health theory emphasizes the role of economic and political structural determinants of individual-level health, focusing on power relations and accountability. Ecosocial theory very explicitly includes consideration of multiple levels corresponding to proposed causal pathways (e.g., individual and neighborhood) and nested hierarchies (e.g., individuals within neighborhoods within cities within states). By definition, theoretical frameworks for investigating environmental health disparities include multiple levels and characteristics of environments and of those residing within them.

Time is an important concept for understanding environmental health disparities. Certain toxicants may have a greater adverse effect during certain ages, as with childhood lead poisoning or fetal exposure to alcohol. These developmental effects have been characterized as “windows of vulnerability” [54, 55]. Further, hazards may flux with time,

such as seasonality of weather and temperature [56]. From a lifecourse perspective, the accumulation of exposures to socioeconomic disadvantage/advantage or socioeconomic characteristics measured at critical time periods (e.g., childhood) could have an important influence on later health outcomes [57]. In addition, relationships between various factors at multiple levels and health are not static. Characteristics of people and places change over time and should be modeled accordingly.

Recent work in social epidemiology discusses the multiple pathways through which socioeconomic characteristics of local places (i.e., neighborhoods) could potentially affect health [10, 58]. One is through the physical environment, including air/water/housing quality, affordable and nutritious food, and safe places to play/exercise. Another pathway is through the social environment, including processes such as social cohesion or the level of mutual trust among neighbors [59]; crime; acceptability of behaviors such as smoking, teen parenting, and adult monitoring of youth; and neighborhood reputation [10]. A third pathway is through the service environment, including fire/police protection, access to health services, transportation, and other social services (e.g., education and job training/placement). Differences in local-level physical, social, and service environments could influence an individual's health through behaviors such as smoking and health care use; psychological stressors such as fear or feeling deprived; hazardous exposures such as pollution, violence, or traffic; and/or opportunities for socioeconomic attainment such as availability of good schools and jobs. Macro-level factors, in turn, directly influence more local-level factors. These multiple pathways have relevance for studying and monitoring environmental health disparities, since they directly impact health or influence vulnerability to environmental hazards.

Similar to studying and monitoring health disparities in general, research on environmental health disparities is rooted in the ethical principle of social justice or equity. Of primary

interest is whether populations that are *a priori* socially disadvantaged in society (by nature of their SEP, race or ethnic group, gender, religious affiliation, etc.) are further—and unacceptably—disadvantaged with regards to their health [60]. Anchoring the study of environmental health disparities in a social justice framework has important operational implications. For example, in investigating disparities, a social justice framework would argue that the reference population (to which other groups are compared) would be the *a priori* most advantaged group (e.g., whites and the highest SEP group) rather than the population average or the group with the lowest risk [60]. Finally, a social/environmental justice approach implies that interventions should be aimed towards health promotion and sustainability rather than remediation only [6].

APPENDIX B: CENSUS DATA AND CENSUS GEOGRAPHY

Census Data

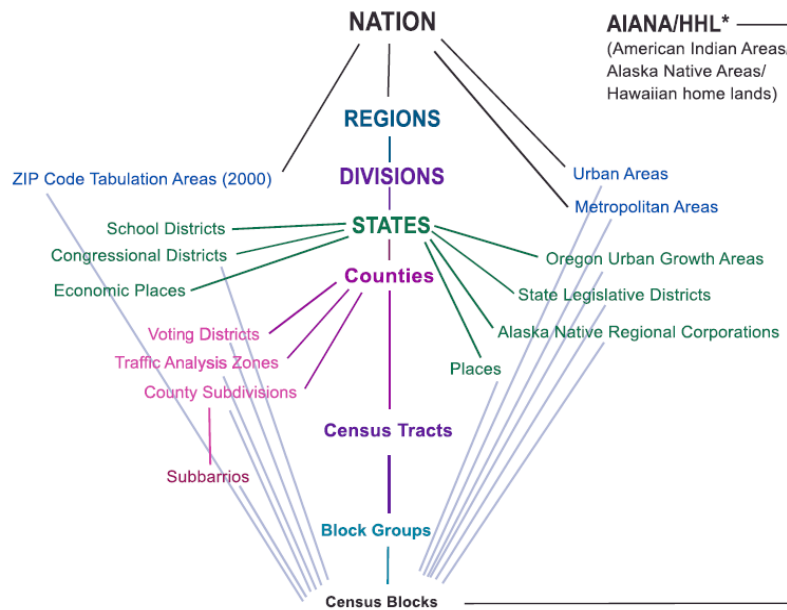
Census data are the most commonly used source for characterizing and defining contexts.

The US Census is generally perceived to be of high quality, given its periodic nature, methodological sophistication, and the relatively small sampling errors. However, one area of concern, especially with regard to racial health disparities, is the well-documented undercount, especially of minority groups. In 2000, estimates of the nationwide undercount range from 0.12 to 1.15%, while the undercount for black males ranged from 2.1 to 7.7%.

The magnitude of the undercount is a less serious consideration than the fact that the undercount varies differentially by race/ethnicity, and home ownership. Various methods for accounting for the undercount have been developed [61, 62]. To overcome the limitation of census data being provided decennially, The American Community Survey will provide comparable data to that available on the US Census on an annual basis for all states, cities, counties, and metropolitan areas. For smaller areas such as census tracts, this survey will release estimates every 5 years [63].

Most Commonly Used Levels

Census blocks are the smallest unit for which the census collects and tabulates data representing approximately 85 individuals (see Figure A1). Visible physical (streets, railroads, and streams) and cultural features (e.g., schools and other buildings) define census blocks. In 1990 the census provided tabulated block data for the entire US, recognizing the utility of these data for small area studies [64].

Figure B1: Standard Hierarchy of Census Geographic Entities

Source: [65]

Block groups are the next level and comprise a cluster of census blocks. Block groups vary in size and generally contain between 600 and 3,000 people (average 1,500). Block group boundaries were originally defined to create population groups that are homogenous with regard to social and economic characteristics [64]. Residential segregation patterns in the US necessitate the use of such small area units for the study of small areas with high minority and/or immigrant populations.

Census tracts, the levels above block groups and composed of block groups, are small relatively permanent geographic areas within counties and comprise between 2,500 and 8,000 residents (average 4,000). Census tracts follow natural boundaries and are designed to be homogeneous with respect to population characteristics and living conditions [64]. Given the relative permanence of census tract boundaries, they are used routinely by

several Federal, state, and local agencies as administrative units for eligibility qualification and resource allocation [66].

Zip codes have commonly been used in health research. Zip codes differ markedly in definition and stability from census tracts and block groups. While census tracts and block groups are delineated to be homogenous units as described previously, zip codes are defined by the US Postal Service (USPS) for efficient mail delivery and can range in size from a single building to large areas that cross state boundaries. To overcome these area discrepancies, the 2000 US Census defined Zip Code Tabulation Areas (ZCTA) mapped to census blocks to replace zip codes. However, the current 5-digit ZCTA area may no longer correspond to the USPS 5-digit zip code area. The use of zip codes is further complicated by extensive modifications in the past 10 years, the creation of new zip codes, and the deletion of existing zip codes. In addition, census will not be providing any linkages between zip codes and ZCTAs. From a theoretical standpoint the use of zip codes as a level in health research is questionable [33].

Counties—and the analogous parish (Louisiana), borough (e.g., Alaska), and county equivalent, hereafter referred to simply as county—are local levels of government, smaller than states but typically larger than cities or towns, with varying degrees of political and legal autonomy depending on the state. Counties provide a useful unit of analysis because they range from completely rural to metropolitan, providing a more representative geography of the US when compared to cities or metropolitan statistical areas. Counties incorporate a wider range of variability in levels of prosperity, demographics, and social and economic infrastructure. Additionally, educational, legal, and political institutions are generally shared within counties.

In contrast to counties, Metropolitan Statistical Areas (MSAs) consist of a large urbanized county or cluster of counties that have a high degree of social and economic integration within that unit. MSAs are often used for programmatic purposes, including allocating Federal funds.

States—and the semi-analogous territories and tribes—are the primary political division in the US and have a large degree of autonomy from the Federal government. This autonomy can challenge Federal environmental laws and regulations, as the extent of implementation is highly variable across states, territories, and tribes.

Variations of Conventional Levels

Given the appropriateness of a specific level and the data availability at that level, one can generate unique geographic and social contexts. For example, census block groups may be combined to create neighborhoods, or census tracts may be combined to create communities. State economic areas, economic sub-regions, and labor market areas can also be considered potential levels of analyses.

Other less commonly used census geographic entities can also be considered [64]. State economic areas (SEAs) are either a single county or a group of counties within a state, defined by economic similarities. An economic sub-region is a group of two or more economically similar counties that cross state lines. Both the SEA and economic sub-region should be reconsidered as important levels between the county and state. Labor market areas (LMAs) are one or more counties defined by commuting-to-work patterns and close economic ties and represent areas within which persons can reside and work within a reasonable distance and change jobs without changing their place of residence [64].

APPENDIX C: INHALATION EXPOSURE TO OUTDOOR AIR POLLUTANTS

Exposure assessment identifies who is exposed, as well as the level and pattern of exposure. Exposure assessment is based on data of the spatial and temporal patterns of air quality and population activity. The nature and complexity of an exposure assessment, including the spatial scale, is a function of the research question, characteristics of the exposure, multiple media sources of the pollutant, measurement methods, and policy.

Table A1 summarizes the elements of various types of exposure assessments. Except for proximity analysis, which provides only qualitative exposure comparisons for disparity analysis, all inhalation exposure assessments require estimates of the spatial pattern of air quality concentrations.

Table C1: Types of Exposure Assessments

Metric	Data Requirements	Analysis Approach	Exposure Accuracy	Resolution/ Extent
Residential proximity to emission sources	(a) Emission magnitudes/locations (b) Residential locations	GIS	Low (qualitative comparisons only)	High
Ambient concentrations at residential locations (monitoring)	(a) Air monitoring (b) Residential locations	GIS	Medium	Low
Ambient concentrations at residential locations (modeling)	(a) Emissions magnitudes/locations (b) Meteorology (c) Residential locations	Air dispersion modeling	Medium	Medium to High
Population exposure (monitoring)	(a) Personal monitoring (b) Residential locations (c) Demographics	Statistics	Low to medium	High
Population exposure (modeling)	(a) Air quality—monitored or modeled (b) Human activity (c) Residential locations	Population exposure modeling	Medium	Medium to High
Individual exposure	Personal monitoring	—	High	Low

Estimating Outdoor Air Pollutant Concentrations

Outdoor air quality can be estimated in several different ways. The estimation procedure directly influences the level at which data are available, with monitoring data available at more aggregate levels such as counties and dispersion modeled data potentially available at the level of individual address.

Air monitoring provides actual measured concentrations at specific locations and times. One of the several limitations of monitoring, however, is that it is expensive. The spatial and temporal extent of the monitoring network is typically small, and/or resolution typically quite coarse. For example, California's South Coast Air Quality Management District recently completed the Multiple Air Toxics Exposure Study (MATES II). To represent air quality for the 15 million residents of Los Angeles, Orange, San Bernardino, and Riverside Counties, concentrations of 32 toxic air pollutants were measured over a year for 24-hour periods every sixth day at 10 fixed-site locations. Air toxics monitoring studies in other regions have measured concentrations at even fewer locations (e.g., eight in Detroit; six in Seattle, Tampa, and Providence; and five in Portland, OR).

Air dispersion modeling studies use data on emissions and meteorology to estimate air quality patterns according to physical principles of atmospheric physics and chemistry. These are generally more cost-effective than monitoring, and thus can be used to estimate air concentrations across larger areas and time periods with finer resolution, potentially to the latitude and longitude of an individual's address. For example, the air dispersion modeling portion of the MATES II study estimated concentrations of air toxics for every 24-hour period over a year at approximately 5,000 locations. US EPA's National Air Toxics Assessment (NATA) national-scale assessment estimates annual average air toxics concentrations for each of the more than 60,000 US census tracts.

Estimating Exposure Concentrations

Similarly, exposure concentrations may be measured directly with personal monitoring, but the cost generally precludes large study samples or monitoring periods longer than a few days. Exposure modeling combines air quality data (monitored or modeled) with data on population activity to estimate population exposure for various populations and demographic subgroups. Simple exposure modeling typically assumes populations are exposed to outdoor concentrations in the vicinity of their residences at all times, similar to the assumption of multilevel models where individuals in a given area share common exposures and experiences. However, more complex modeling uses population activity data from diary studies to account for people's movements among indoor and outdoor microenvironments and geographic locations. The NATA national-scale assessment uses such a model to estimate exposure concentrations for 10 age-gender groups in each US census tract.

Characteristics of the Pollutant

Steeper concentration gradients require finer spatial resolution (e.g., census tracts or block groups) to accurately represent variations in exposure concentrations. Steeper gradients are expected for primary pollutants (i.e., those emitted directly from sources) than for secondary pollutants (i.e., those formed on the atmosphere from chemical transformation of precursor compounds, such as tropospheric ozone). Steeper gradients are also expected for large isolated emission sources (e.g., large power plants and large industrial facilities) than for sources that are more dispersed (e.g., on-road motor vehicles). However, even for onroad motor vehicles there are steep concentration gradients within a couple hundred meters of a major roadway, so the contextual setting of the study population is also an important consideration.

Air Quality Research Goals

Air Quality Standards

The determination of whether an air quality standard has been violated requires the estimation of the maximum concentration in the modeling domain. When an air quality study is focused on a single, dominant emission source (e.g., a large isolated stationary emission source or a heavily trafficked, isolated roadway or intersection), concentration measurements may be made with a dense network of monitors spread over a very limited area. However, due to their limited spatial scope, such studies are not likely to be useful for health disparity analysis. For primary pollutants (i.e., steep gradients near emission sources), air dispersion modeling for maximum concentration determination is often done in a tiered manner, using coarse spatial resolution for the first tier to find the general area of highest concentrations and then using finer resolution on a subset of the domain in the second tier. The modeling receptors may be arranged in a gridded pattern or be associated with census subdivisions (e.g., internal points of tracts, block groups, and blocks).

Population Exposure Assessment

The goal of a population exposure analysis may be to find the average exposure concentrations for various populations, stratified by geography and/or demography. In this case the choice of spatial resolution may depend on the resolution of the populations of interest, as well as the spatial variability of the air quality concentrations. The census tract has been demonstrated generally to be an appropriate level to approximate the average exposure concentration for constituent ethnic groups adequately. Exceptions remain for highly segregated areas such as Harris, TX. In such cases census blocks or block groups

may be more appropriate; however, some desired demographic details may not be available for the more spatially resolved data (see Table A2 and A3 below).

Table C 2: Percentage of Ethnic Group Members that Reside in US Census Blocks in which the Ethnic Group Fraction Is within 0.10 of the Ethnic Group Fraction of the Tract within which it Resides

COUNTY	ETHNIC GROUP			
	White	Hispanic	Black	Asian
Los Angeles	74%	68%	67%	73%
Orange, CA	80%	67%	96%	80%
Brooklyn	81%	77%	85%	80%
Manhattan	88%	81%	86%	78%
Lorraine, OH (Cleveland)	88%	77%	83%	90%
Harris, TX (Houston)	60%	53%	58%	73%

Source: 1990 US Census data

Table C 3: Ethnic Composition of Six Selected US Counties

COUNTY	TOTAL POP.	ETHNIC GROUP			
		White	Hispanic	Black	Asian
Los Angeles	8,863,164	41%	38%	11%	10%
Orange, CA	2,410,556	64%	23%	2%	10%
Brooklyn	2,300,664	40%	20%	35%	5%
Manhattan	1,487,536	49%	26%	18%	7%
Lorraine, OH (Cleveland)	1,412,140	72%	2%	25%	1%
Harris, TX (Houston)	2,818,199	54%	23%	19%	4%

1990 US Census data

Alternatively, the goal of a population exposure analysis may be to find the number of persons exposed above a given threshold. In this case, for steep concentration gradients, a finer spatial resolution may be required so that spatial averaging does not mask exceedances of the threshold, e.g., census block rather than tract.

Multiple Media

Multiple sources of air pollution data are an important consideration in determining the appropriate levels in the analytic design. Outdoor air pollutants are emitted by a variety of sources: major stationary sources, small industrial sources, commercial facilities, residences, on-road motor vehicles, and non-road mobile equipment. Residential proximity to the dominant source type for a given pollutant is often used as a surrogate measure for exposure to that pollutant when air quality concentration data are unavailable. Proximity to the source type can also be used as a surrogate for estimating spatial patterns of concentrations within the geographic subdivisions for which concentration data are available. For example, it is estimated that about 60% of CO exposure typically results from on-road vehicle emissions. If CO concentration estimates were available at the tract level, variations within the tract (e.g., at the block level) could be estimated based on traffic patterns within the tract.

Data Sources

US EPA's NATA national-scale assessment: Modeling estimates of 1996 annual average census tract ambient concentrations and exposure concentrations of 33 air toxics. Modeling estimates of 1999 annual average census tract ambient concentrations and exposure concentrations are scheduled to be reported in Spring 2005.

South Coast Air Quality Management District (SCAQMD) MATESII study: Daily and annual average (1998-1999) air toxics monitoring (10 locations) and modeling estimates (2 km spacing) for California's South Coast Air Quality Management District (Los Angeles, Orange, San Bernardino, and Riverside Counties).

Portland, OR Air Toxics Assessment (PATA): Daily and annual average (1999-2000) air toxics monitoring (five locations) and modeling estimates (block group and tract resolution).

SCAQMD Air Quality Management Plan: Episodic concentrations of tropospheric ozone (secondary), primary and secondary particles, and carbon monoxide (5 km spacing) for California's South Coast Air Quality Management District (Los Angeles, Orange, San Bernardino, and Riverside Counties).

Tropospheric ozone monitoring network: hourly concentration measurements at approximately 1,200 locations in the US.

Particulate matter (PM) network: 24-hour average concentration measurements every sixth day at approximately 1,000 locations in the US for PM₁₀ and approximately 1,000 locations for PM_{2.5}.

Policy Level

For criteria pollutants (ozone, carbon monoxide, nitrogen dioxide, sulfur dioxide, particulate matter, and lead) US EPA sets air quality standards. California has also set somewhat different air quality standards, generally stricter than US EPA's. States, and in some cases local air quality management districts, are responsible for attaining the standards within their jurisdictions by regulating emission sources. US EPA also promulgates emission regulations directly for some types of sources, such as onroad motor vehicles and some types of nonroad mobile equipment. US EPA also has promulgated control rules for some sources of non-criteria pollutants, i.e., air toxics.

APPENDIX D: AMBIENT AND DRINKING WATER QUALITY

Characteristics of the Exposure

Exposure to environmental agents in water can occur through a variety of pathways, including: dermal contact, incidental ingestion, and inhalation during swimming in rivers, lakes, and oceans; ingestion of fish and shellfish from contaminated waters; direct ingestion of drinking water; indirect ingestion of water used to process foods; and inhalation of radon and other contaminants entering homes from groundwater or during showering.

The sources of the environmental agents also vary widely and include direct disposal of wastewaters, runoff from farms, groundwater contamination from land disposal, air emissions that subsequently deposit to surface waters, naturally occurring materials (e.g., geological arsenic and radiological deposits that leach into groundwater), and poorly constructed or maintained drinking water treatment and distribution systems.

One of the more studied water pathways—direct ingestion of drinking water—can require a variety of spatial and other levels of analysis, depending on the populations and contaminants being studied, which in turn are related to the physical and regulatory framework of drinking water systems. Briefly, US EPA regulates public water systems (PWS's), the US Food and Drug Administration (FDA) regulates bottled water, and state and local authorities—or no one—regulates very small or private supplies (i.e., private wells and systems that are too small to be considered PWS's). Each of these sources of drinking water can have widely differing—or no—standards and, thus, widely differing exposures and risks.

Water Systems

US EPA's definition of a public water system (PWS) states that it is a system for the provision to the public of water for human consumption through pipes or other constructed conveyances, if such a system has at least 15 service connections or regularly serves at least 25 individuals at least 60 days out of the year. A PWS can be one of two basic types:

- Community water systems (CWS's) serve at least 15 service connections or 25 people year-round in their primary residences. Most residences, including homes, apartments, and condominiums in cities, small towns, and mobile home parks, are served by CWS's.
- Non-community water systems (NCWS's) are public water systems that serve the public but do not serve the same people year-round. There are two types of NCWS's:
 - *Non-transient non-community systems* (NTNCWS's) serve at least 25 of the same persons over 6 months per year (e.g., schools or factories that have their own water source).
 - *Transient non-community systems* (TNCWS's) serve at least 25 persons (but not the same 25) over 6 months per year (e.g., campgrounds or highway rest stops that have their own water source).

While the majority of people (about 65%) obtain most of their drinking water from PWS's, the rest are about evenly divided between drinking primarily bottled water and water from private wells [67, 68]. Furthermore, residents of rural or farm areas are more likely to drink private well water than municipal or bottled water. About half of drinking water in the US is from surface water, and half is from ground water, although often several sources of water are blended for a given PWS.

The source of drinking water can have a significant effect on the type, concentration, and frequency of exposure to environmental agents. Given the variety of drinking water sources in some areas, therefore, exposure in these areas also can vary widely. In some cases, individuals in the same neighborhood or even the same household can experience different exposures if, for example, some individuals rely solely on municipal tap water (e.g., from a CWS) for their drinking water, some rely on bottled water, and some rely on water primarily from work (e.g., a NTNCWS). In a study of inter- and intra-ethnic variation in water intake among Tucson residents, for example, Hispanics reported differences such as much higher rates of bottled water consumption than did non-Hispanic whites [3].

Rural areas, where individual households often obtain their water from private wells, can experience very different exposures even when the wells are fairly close to each other because of local effects of landfills or other sources of contamination but also because the wells might be drilled to different depths and, thus, drawing from different aquifers. Exposure can also vary within a larger CWS. For example, contamination may occur close to the tap (e.g., lead in the home's faucet or pipes) rather than at the water source.

Monitoring Research Goals

The estimation of risk or determination of whether a drinking water standard has been violated usually requires the sampling and analysis of water samples, often at the PWS level prior to distribution to homes (i.e., before the water enters the pipes to and in the homes) and sometimes at the water's source (e.g., river, aquifer). At times, sampling is conducted of contaminated soils or other material (e.g., landfill contents) "upstream" of the water, and then modeling is conducted to estimate concentrations in the ingested water. Sometimes additional sampling or modeling of changes in concentrations after the water leaves the PWS or other system is required.

PWS's and bottlers are required to conduct monitoring on a regular basis, though, as described above, the standards vary depending on the type of water supply. Also, sampling of water at the tap is not conducted regularly, and, thus, modeling often is needed.

In one example of modeling of trihalomethane (THM) concentrations from multiple sources of water in a PWS, a statistical model was constructed using sparse routinely collected THM measurements to obtain quarterly estimates of mean THM concentrations. The THM measurements were modeled using a Bayesian hierarchical mixture model, taking into account heterogeneity in THM concentrations between water originating from different source types, quarterly variation in THM concentrations due to factors such as precursor concentrations and the time the water has spent in the distribution system, and uncertainty in the true value of undetected and rounded measurements [69].

Using water monitoring or modeling data with demographic data at the census tract level can be difficult. For example, in a study of problems associated with collecting drinking water quality data for community studies, the task of evaluating water quality for each census tract was complicated by the fact that single census tracts were served by more than one system; each system usually had more than one well; and single wells had several episodes of testing for various contaminants [42]. This is a classic problem that the multilevel modeling approach described in “multiple contexts at the same level” can potentially address.

Other monitoring focuses more on the drinking water source. For example, McLaughlin et al. (2001) [70] note how ambient water (not all of which is used as a source of drinking water) in predominately black counties received more than twice the mass of chemicals released to water per square kilometer than all counties.

Actual health outcome attributable to drinking water, as with other public health data, can be collected and aggregated at a variety of different levels, including individuals, census blocks,

census tracts, counties, and health districts, depending on whether the data are obtained from national, state, or local disease surveillance registries or from specific surveys. Data used for Los Angeles County in 2000 and New York County in 2001 demonstrate the use of a variety of different geographic units—primarily health districts and neighborhoods, respectively—to demonstrate that Hispanics have higher rates of giardiasis and cryptosporidiosis, respectively, than other ethnic groups [71].

Monitoring data are available from a variety of sources, including the American Water Works Association (AWWA), US EPA's Safe Drinking Water Information System (SDWIS), and state and local utilities reports. Such data usually are provided at the water supply level, which often equate to counties but also can equate to multiple counties or multiple units within counties. Thus, census tracts often combine to form congruent areas with water supply areas, yet a single census tract can also contain more than one water supply area or parts of areas.